

Optional Step 4:
Formatting GWAS data using sumstats

sumstats takes 7 key arguments:

- 1. files:** The name of the summary statistics files. This should be the same as the name of the files used for the munge function in Step 1 above **and the files should be in the same listed order used for the ldsc function in step 2.** ★
- 2. ref:** The reference file used to calculate SNP variance across traits. We use 1000 genomes phase 3 that is available for download here: <https://utexas.box.com/s/vkd36n197m8klbaio3yzoxsee6sxo11v>.
- 3. trait.names:** The names of the traits in the order that they are listed for the files.
- 4. se.logit:** Whether the SEs are on a logistic scale.
- 5. OLS:** Whether the phenotype was a continuous outcome analyzed using an observed least square (OLS; i.e., linear) estimator. Not used here since all binary traits
- 6. linprob:** Whether the phenotype was an outcome for which there are only Z-statistics in the summary statistics file -or- it was a dichotomous outcome analyzed using an OLS estimator.
- 7. N:** A user provided sample size. When backing out a beta using the linprob argument this requires the **sum of effective sample sizes**.

Flowchart on GitHub to help you figure out arguments for sumstats

The linear probability model argument for alcohol use

sumstats needs a beta and SE to perform analyses, but sometimes the summary data only has a Z-statistic (or it was a binary outcome analyzed as continuous).

In these cases, `linprob` is set to `TRUE` for that trait and the sum of effective sample size is needed

	CHR	SNP	BP	A1	A2	Z	P	Weight
1	1	rs10799799	23973601	T	G	-0.190	0.8495	21857.28
2	1	rs200328701	57729103	CT	C	-0.278	0.7807	18324.09
3	1	rs75245025	151716030	C	G	0.710	0.4779	22791.88
4	1	rs146090341	211396305	T	G	-0.092	0.9265	18999.23
5	1	rs188551793	215407827	A	T	0.321	0.7482	11611.03
6	1	rs2490387	237279677	T	C	-1.216	0.2239	20769.08

Examine end of full GWAS .log file

After merging across all summary statistics using listwise deletion, performing QC, and merging with the reference file, there are **5301314** SNPs left in the final multivariate summary statistics file

Running sumstats for all files took **9 minutes and 4.62626385688782 seconds**

Please check the log file MDD_PTSD_ANX_ALCH_sumstats.log to ensure that all columns were interpreted correctly and no warnings were issued for any of the summary statistics files.